Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 51-29 (2018) 400–405

Pose Estimation Considering an Uncertainty Model of Stereo Vision for In-Water Ship Hull Inspection

Dongha Chung and Jinwhan Kim

Department of Mechanical Engineering, KAIST, Daejeon, Korea
(e-mail: chungdongha@kaist.ac.kr, jinwhan@kaist.ac.kr)

Abstract: For visual inspection of underwater structures by using an autonomous underwater vehicle (AUV), the relative pose between the AUV and the underwater structure needs to be estimated accurately. In this paper, a vision based pose estimation method for in-water ship hull inspection using 3D point cloud data provided by a stereo vision system is proposed. In particular, a systematic procedure to evaluate the uncertainty of feature point location on the obtained stereo images is formulated and applied. An experimental result is presented to demonstrate the practical feasibility of the proposed approach.

Keywords: Stereo vision, Autonomous underwater vehicle, Underwater hull inspection

1. INTRODUCTION

Periodical in-water ship hull inspection is important for early detection of any surface degradation, damage and biofouling to maintain the safety and efficiency of various ship operations. Generally, such tasks are performed by human divers, which has apparent limitations regarding the efficiency and safety involved.

The use of AUVs adopting the underwater visual mapping method (Hong et al., 2016; Elibol et al., 2014) can be an useful means for the ship-hull inspection. Figure 1 shows an illustration of visual mapping method using an AUV for hull inspection.

For in-water visual inspection of the ship hulls using an AUV, the relative pose estimation of the AUV toward the hull surface is crucial. Sonar sensors were used for achieving the relative pose with respect to the target surface for such survey missions (Kazmi et al., 2009; Vaganay et al., 2005). In Negahdaripour and Firoozfam (2006), a stereo vision system was proposed to achieve the relative pose using a stereo matching process based on the difference in pixel intensity. In Chung et al. (2017), a stereo vision system based on the Speeded Up Robust Features (SURF) (Bay et al., 2006) was proposed to estimate the relative pose of the camera toward an underwater planar surface. However, the method was based on a deterministic relative pose measure.

In this paper, a relative pose estimation method considering a systematic uncertainty model of stereo vision is introduced. For this, the following procedure is proposed. Firstly, the feature point uncertainty in the image plane using the characteristics of SURF feature detection is estimated. Secondly, the uncertainty of the 3D point location is propagated through a transformation function which maps the 2D disparity map to the 3D point cloud. Thirdly, a least square method is employed to fit a plane into the point cloud using the Mahalanobis distances (Mahalanobis, 1936). The uncertainty of the plane parameters is then estimated by Taylor approximation. Finally, the relative pose and its uncertainty are estimated using the geometrical relationship between the plane and relative pose.

2. SPATIAL COGNITION WITH STEREO VISION

For precise matching of stereo correspondences, a camera calibration procedure including image undistortion and stereo rectification is applied, and the contrast limited adaptive histogram equalization (CLAHE) (Yadav et al., 2014) is used to compensate the non-uniform illumination effects on the image.

2.1 Stereo matching with SURF operator

The absorption and scattering of light in an underwater environment limit the performance of computer vision algorithms. The conventional block matching algorithm for finding the stereo correspondence often fails to provide correct matching results due to the characteristics of the target surface which is monotonic and texture-less, and the blurring effect which significantly increases with the distance.

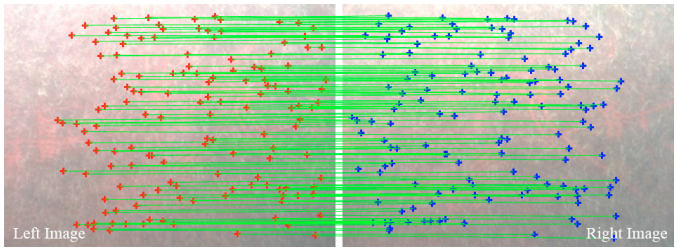
To overcome this problem, feature extraction and matching algorithms such as the Harris corner (Harris and Stephens, 1988), scale-invariant feature transform (SIFT) (Lowe, 2004), and SURF operators can be employed. The SIFT operator is widely used in visual-feature-based applications because of its high performance. However, the computational cost of the SIFT operator is relatively high, which often limits its application to online estimation. In this study, the SURF operator is employed to achieve improved computational efficiency.

Once the SURF features are detected in both left and right images, the matching process for stereo correspondences is

Fig. 1. Illustration of underwater hull inspection using an AUV.

10.1016/j.ifacol.2018.09.454

(a) Original stereo images.



(b) Matching result of SURF feature points.

Fig. 2. An example of matching SURF feature points for stereo images.

performed considering the following constraints. Firstly, with consideration of the geometrical relationship between the left and right cameras, each stereo correspondence should lie near the epipolar line. To simplify the epipolar line, the stereo images are rectified so that the epipolar line becomes a horizontal line. Secondly, the features with similar scale values are matched since the distances from a 3D point to the left and right cameras are similar. Thirdly, as no 3D point can be located at the back of the stereo camera, the stereo correspondences are matched only when their disparity values are positive (i.e., the feature point on the left image should always be located on the right side of the feature point location on the right image). With the feature points following these constraints, the stereo correspondences are matched between the features with least difference in descriptor vectors. Figure 2 shows an example of the result of stereo matching based on these conditions.

2.2 3D point cloud extraction and uncertainty propagation

A 3D point location is extracted from the matched results using the geometrical relationship between the 3D point and its projection into rectified image planes as follows:

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \frac{b}{d_i} \begin{bmatrix} 0 & 0 & f \\ 1 & 0 & -x^l_{pcc} \\ 0 & 1 & -y^l_{pcc} \end{bmatrix} \begin{bmatrix} x^l_{pi} \\ y^l_{pi} \\ 1 \end{bmatrix} \tag{1}$$

where (X_i, Y_i, Z_i) is the 3D point location expressed in left camera reference frame for the i$^{\text{th}}$ stereo correspondence pair, (x^l_{pi}, y^l_{pi}) is the feature point location on the left image, (x^l_{pcc}, y^l_{pcc}) is the center of the left camera, f is the focal length of the left camera,b is the baseline length of the stereo camera, and d_i is the disparity value of the stereo correspondence pair ($d_i = x^l_{pi} - x^r_{pi}$, where x^r_{pi} is the x coordinate of feature point location of the right side of the image).

The uncertainty of the 3D point location is propagated considering the uncertainty of the corresponding SURF feature location in the image space. The location uncertainty (covariance matrix) of the SURF feature point was formulated in Zeisl et al. (2009). Assuming that the underwater images are blurred and monotonic, the covariance of the i$^{\text{th}}$ feature point location can be approximated by (2), where σ_{pi} is the scale of that feature point.

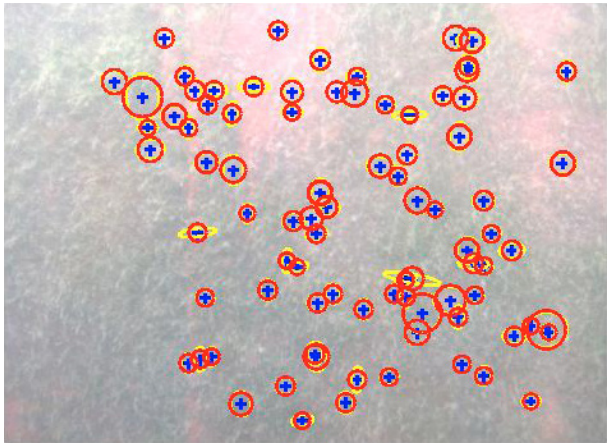


Fig. 3. An example of location uncertainties of feature points in an underwater scene.

$$\Sigma_{\mathbf{X}pi} = \sigma^2_{pi} I_{2\times 2} \tag{2}$$

Figure 3 shows the uncertainty boundaries of the feature points based on the covariance matrix calculated by Zeisl et al. (2009) (yellow ellipsoids) and (2) (red ellipsoids). Using such an approximation, the covariance matrix, $\Sigma_{\mathbf{D}i}$, for the parameters $(x^l_{pi}, y^l_{pi}, d_i)$ is expressed by:

$$\Sigma_{\mathbf{D}i} = \begin{bmatrix} {\sigma^l_{pi}}^2 & 0 & {\sigma^l_{pi}}^2 \\ 0 & {\sigma^l_{pi}}^2 & 0 \\ {\sigma^l_{pi}}^2 & 0 & {\sigma^l_{pi}}^2 + {\sigma^r_{pi}}^2 \end{bmatrix} \tag{3}$$

where σ^l_{pi} and σ^r_{pi} are the scale values of the feature points lying on the left and right images.

The covariance of the 3D point location can be propagated from (3) using the relationship between the 3D point coordinates and the parameters $(x^l_{pi}, y^l_{pi}, d_i)$, as shown in (1). Equation (1) is nonlinear, and the Taylor's first order approximation is employed to express the relationship between the perturbation of the 3D point location and the perturbation of the feature point location on the image plane and disparity as follows:

$$\begin{bmatrix} \Delta X_i \\ \Delta Y_i \\ \Delta Z_i \end{bmatrix} \approx J_i \begin{bmatrix} \Delta x^l_{pi} \\ \Delta x^l_{pi} \\ \Delta d_i \end{bmatrix} \tag{4}$$

where J_i is the Jacobian of the transformation matrix.

The above approximation is valid when the higher order terms are proportional to $(\Delta d_i^n / (n! \times d_i^{n+1}))$ where n is an integer greater than one. This indicates that the farther the 3D point is located from the camera, the smaller is its disparity value; thus, the approximation may induce a larger error. When the target surface is close to the camera ($\sim$ 2 m), such an approximation is found to be feasible.

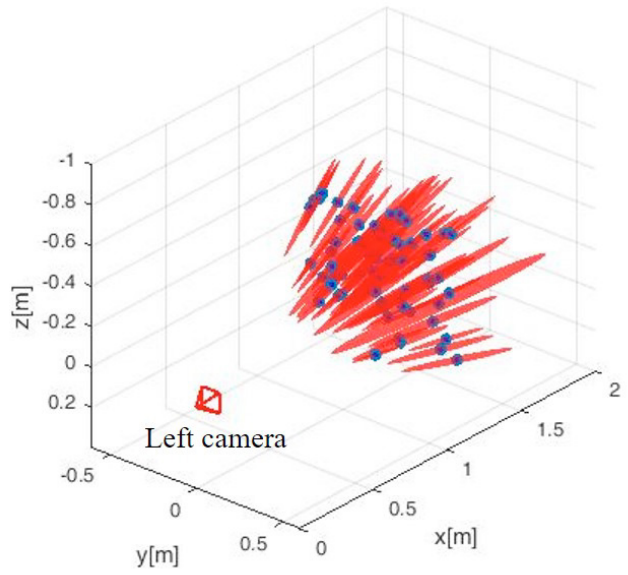


Fig. 4. An example of estimated 3D point locations (blue points) and their uncertainty boundaries (red ellipsoids).

Using the relationship between the perturbation of the 3D point location and the parameters as in (4), the covariance matrix of the 3D point location $\Sigma_{\mathbf{P}i}$ is calculated as follows:

$$\Sigma_{\mathbf{P}i} = J_i \Sigma_{\mathbf{D}i} J_i^T. \tag{5}$$

Figure 4 displays the extracted point cloud and the location uncertainty of each point inside the point cloud. Please note that the uncertainty ellipsoids spread in the direction of the camera center to the points, as the Y and Z coordinates are proportional to the X coordinate. In addition, as the disparity value is located in the denominator of (1), its uncertainty value induces a large variance in the depth (X) direction. This propagation method uses the first order approximation of the transformation function; therefore, this may not be valid for the case when the disparity value is small (i.e., 3D point is far away from the camera), as higher order terms may be significant. In the case of the proposed system, the higher order terms are assumed to be negligible as this system is operated near the target surface.

3. RELATIVE POSE ESTIMATION

By assuming that the local hull surface is planar, the relative pose to the surface can be estimated by fitting a plane onto the point cloud which is gathered using the method introduced in Section 2.

3.1 RANSAC with plane model

Most plane fitting algorithms for point cloud data use the least squares method, which estimates the plane to have a minimum sum of the squared distance from each point. However, each point can affect the result of this method, leading to a wrong estimation in the case of some mismatched points. For the point cloud gathered in Section 2, there may exist some mismatched points. To eliminate such mismatched points, the RANSAC algorithm (Fischler and Bolles, 1981) is employed, with a plane model as its base function. With such a method, the points that follow the dominant plane model (inlier points) can be extracted and used for plane fitting with the least squares method.

3.2 Least square method for plane fitting

With the points classified as inliers by the RANSAC algorithm, the least squares method to estimate the plane model is used. The sum of the squared Mahalanobis distance from each point to the plane is defined as the cost function using the estimated location uncertainty of the point set. as follows:

$$S_{LSQ} = \sum_{i=1}^{N} (\mathbf{P}_i - \mathbf{p}_i)^T \Sigma_{\mathbf{P}i} (\mathbf{P}_i - \mathbf{p}_i). \tag{6}$$

For the i$^{\text{th}}$ point $\mathbf{P}_i$, the point $\mathbf{p}_i$ is set to be a projection of $\mathbf{P}_i$ into the plane; thus, the Mahalanobis distance from the point to the plane is $(\mathbf{P}_i - \mathbf{p}_i)^T \Sigma_{\mathbf{P}i} (\mathbf{P}_i - \mathbf{p}_i)$. Although the plane equation is expressed as $\alpha x + \beta y + \gamma z + \delta = 0$, with the plane parameters of (α, β, γ, and δ), the degree of freedom is three, as it is invariant to the scale of the parameters. Thus, the plane model can be set as $x + ay + bz + c = 0$, with plane parameters (a, b, and c). By solving the nonlinear equations of $\frac{\partial S_{LSQ}}{\partial a} = 0$, $\frac{\partial S_{LSQ}}{\partial b} = 0$, and $\frac{\partial S_{LSQ}}{\partial c} = 0$, the plane parameters that satisfy the least squares solution are obtained.

3.3 Uncertainty propagation of plane parameters

The uncertainty propagation of the plane parameters from a point cloud, which has a different uncertainty value for each point, is done by evaluating the effect of perturbation of an individual point location on the plane parameters. For this procedure, a function $g(\mathscr{P}, \mathscr{A})$ is set as the partial derivative of the cost function with respect to the plane parameters as follows:

$$g(\mathscr{P}, \mathscr{A}) = \frac{\partial S_{LSQ}}{\partial \mathscr{A}} = \begin{bmatrix} \frac{\partial S_{LSQ}}{\partial a} \\ \frac{\partial S_{LSQ}}{\partial b} \\ \frac{\partial S_{LSQ}}{\partial c} \end{bmatrix} \tag{7}$$

where $\mathscr{P}$ is a $3N \times 1$ vector containing all the coordinates of points as $(\mathbf{P}_1, \mathbf{P}_2, ..., \mathbf{P}_i, ..., \mathbf{P}_N)$, and $\mathscr{A}$ is a 3×1 vector containing the plane parameters (a, b, c). From the result of the previous section, $g(\mathscr{P}, \mathscr{A}) = 0$ where $\mathscr{A}$ is the least squares solution. The covariance matrix of $\mathscr{A}$ can be calculated using the propagation method in Haralick (2000) as follows:

$$\Sigma_{\mathscr{A}} = (\frac{\partial g}{\partial \mathscr{A}})^{-1} \frac{\partial g}{\partial \mathscr{P}} \Sigma_{\mathscr{P}} (\frac{\partial g}{\partial \mathscr{P}})^T (\frac{\partial g}{\partial \mathscr{A}})^{-T}. \tag{8}$$

$\mathscr{P}$ is the vector containing the point coordinates in sequence and $\Sigma_{\mathscr{P}}$ is a block diagonal matrix in which each block is the covariance matrix of its corresponding point.

3.4 Conversion of the plane model to relative pose

By using the geometrical relationship as shown in Fig. 5, the plane parameters (a, b, c) of the plane ($x + ay + bz + c = 0$) can be converted into relative pose (relative yaw (ψ_{rel}), pitch (θ_{rel}), and standoff distances (d_{rel})) as follows:

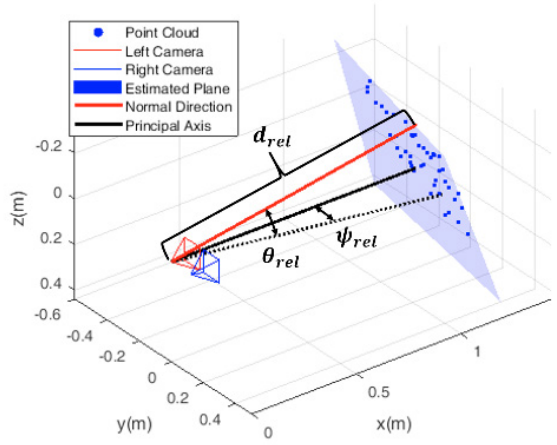


Fig. 5. The Relationship between the estimated plane (blue plane) and the relative pose.

$$\begin{bmatrix} \psi_{rel} \\ \theta_{rel} \\ d_{rel} \end{bmatrix} = \begin{bmatrix} atan2(a,1) \\ -atan2(b,\sqrt{1+a^2}) \\ \frac{|c|}{\sqrt{1+a^2+b^2}} \end{bmatrix}. \tag{9}$$

The covariance of $(\psi_{rel}, \theta_{rel}, d_{rel})$ can be propagated using the Taylor's first approximation as follows:

$$\Sigma_S = J \Sigma_{\mathscr{A}} J^T \tag{10}$$

where J is the Jacobian of the transformation matrix as:

$$J = \begin{bmatrix} \frac{\partial \psi_{rel}}{\partial a} & \frac{\partial \psi_{rel}}{\partial b} & \frac{\partial \psi_{rel}}{\partial c} \\ \frac{\partial \theta_{rel}}{\partial a} & \frac{\partial \theta_{rel}}{\partial b} & \frac{\partial \theta_{rel}}{\partial c} \\ \frac{\partial d_{rel}}{\partial a} & \frac{\partial d_{rel}}{\partial b} & \frac{\partial d_{rel}}{\partial c} \end{bmatrix}. \tag{11}$$

4. EXPERIMENTAL VALIDATION

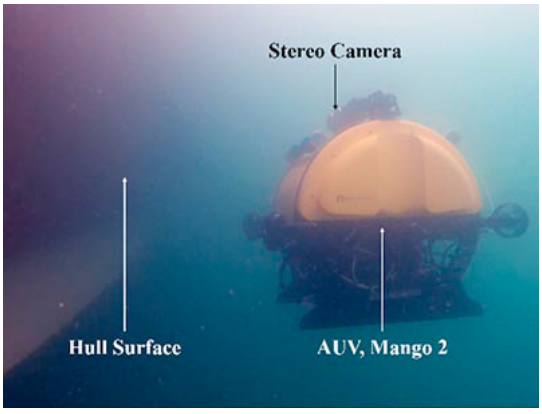
The proposed algorithm for determining the relative pose and its uncertainty was applied to the dataset obtained around a full-scale ship (Fig. 6a) using an AUV (Fig. 6b). The detailed vehicle specifications of the AUV are described in (Hong et al., 2017). The dataset covers the hull surface area with a width of 19m and a height of 2m which contains the planar region in the mid-ship and the sloped region (deadrise) in the fore part of the ship. The heading angles of the vehicle and the stereo camera were controlled to maintain its initial heading. Therefore, the dataset contains various relative poses toward the hull surface. The ground-truth data of the relative pose is generated by manually matching the stereo pairs, since no other underwater positioning system was available.

4.1 Experimental result

The algorithm was tested across various situations, as the status of the hull surface varied along the ship hull surface. Figure 7 displays some examples of the hull surface scenes and the result of the proposed algorithm applied to them.

The measurement error varied depending on the status of the image. For example, for the images shown in Fig. 7, the measurement errors are negligible for the feature-rich regions, where the SURF features are well-distributed and their scales are small. However, the measurement errors are larger for the feature-poor and sloped regions, where the SURF features are relatively concentrated, and their scales are large. The measurement uncertainty varies depending on such conditions (feature distributions and feature scales), the pose estimation error is affected by the uncertainty defined by the covariance matrix, as shown in Table 1 and Table 2.

(a)



(b)

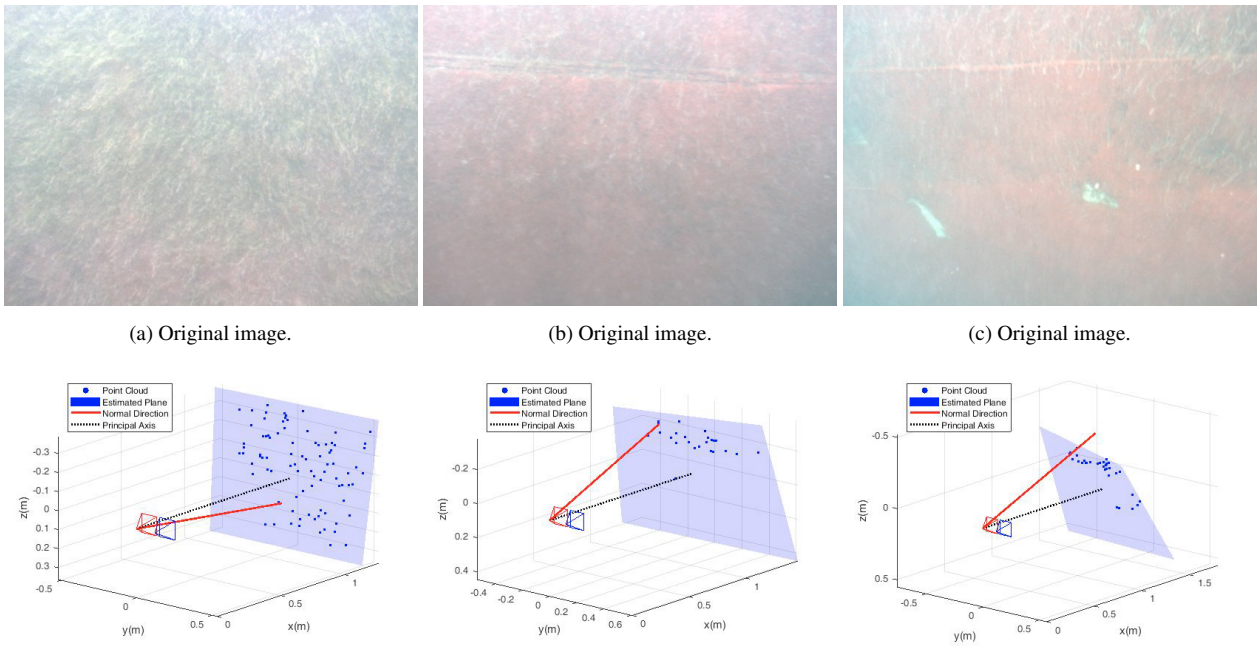
Fig. 6. The target ship (KIOST R/V Onnuri) (a) and an underwater scene of the AUV inspecting the hull surface (b).

Table 1. Pose estimation error

Image Condition	Yaw error [deg]	Pitch error [deg]	Distance error [m]
Feature rich	0.241	0.732	0.036
Feature poor	15.284	5.470	0.150
Sloped surface	3.525	1.913	0.238

The evaluation is done by applying the algorithm to 1388 stereo images taken on various locations of the hull surface. Since the uncertainties of the measurement values differ depending on the image characteristics, the errors are normalized by utilizing their covariance matrices. For the ith measurement covariance matrix $\Sigma_{\mathbf{M}i}$, the lower triangular matrix L_i can be calculated by the Cholesky decomposition as follows:

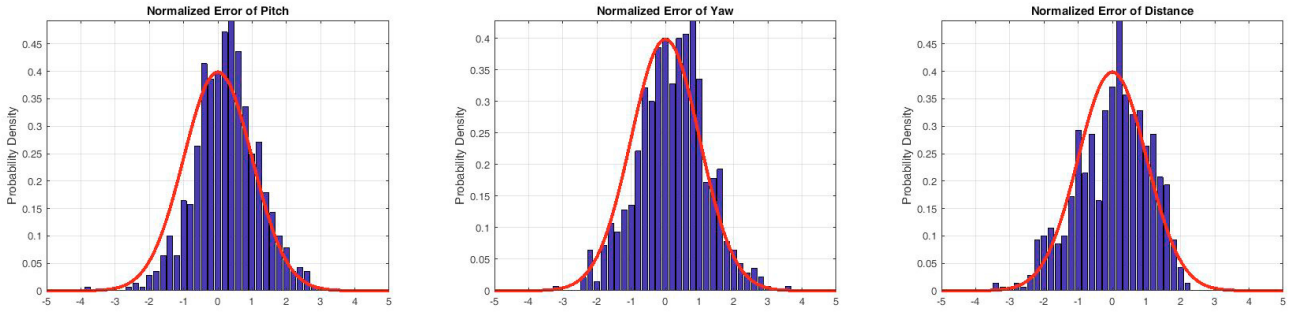
$$\begin{aligned} \Sigma_{\mathbf{M}i} &= L_i L_i^T \\ {L_i}^{-1} \Sigma_{\mathbf{M}i} {L_i}^{-T} &= I. \end{aligned} \tag{12}$$



(a) Original image. (b) Original image. (c) Original image.

(d) Measured result. (e) Measured result. (f) Measured result.

Fig. 7. Performance evaluation on different underwater hull surface scenes: feature rich region (1st column), feature poor region (2nd column), and sloped region (3th column).



(a) Normalized pitch error. (b) Normalized yaw error. (c) Normalized distance error.

Fig. 8. Histogram of normalized errors of measured relative pose values.

As the covariance matrix of the linear transformation of $y = Ax$ can be propagated as $\Sigma_y = A\Sigma_x A^T$, the i$^{\text{th}}$ error, ε_i, can be transformed by multiplying the transformation matrix as follows:

Table 2. Measurement covariance matrix

Image Condition	Covariance Matrix [$\times 10^{-3}$]
Feature rich	$\begin{bmatrix} 0.396 & 0.038 & 0.076 \\ 0.038 & 0.670 & 0.126 \\ 0.076 & 0.126 & 0.052 \end{bmatrix}$
Feature poor	$\begin{bmatrix} 2.675 & 0.127 & 0.856 \\ 0.127 & 1.465 & -0.004 \\ 0.856 & -0.004 & 0.388 \end{bmatrix}$
Sloped surface	$\begin{bmatrix} 1.238 & 0.058 & 0.229 \\ 0.058 & 1.193 & 0.148 \\ 0.229 & 0.148 & 0.163 \end{bmatrix}$

$$\varepsilon_i' = {L_i}^{-1}\varepsilon_i. \tag{13}$$

The covariance matrix at each step is normalized as an identity matrix, so the normalized errors ($\varepsilon_\square'$) are expected to follow the standard multivariate Gaussian distribution. The performance and validity of the proposed uncertainty modeling method are shown in Fig. 8. It can be seen that the of the distributions of normalized errors (blue bars) follow the standard Gaussian distribution (red curves) reasonably well.

The qualitative result is shown in 3D image patches using the estimated relative poses and the camera position (see Fig. 9). Figure 9a shows the view from the outer part of the hull surface, whereas Fig. 9b shows the view from the inner part of the surface. To compare the result with an actual ship model, the point cloud data of the actual ship's 3D model are also plotted as blue dots. The negative and positive y directions represent the fore and aft directions of the ship, respectively. Figure 9c shows the frontal view of the resultant 3D patches. The plane patches agree with the ship's actual 3D model (Fig. 9a and Fig. 9b) and show a reasonable mapping result (Fig. 9c).

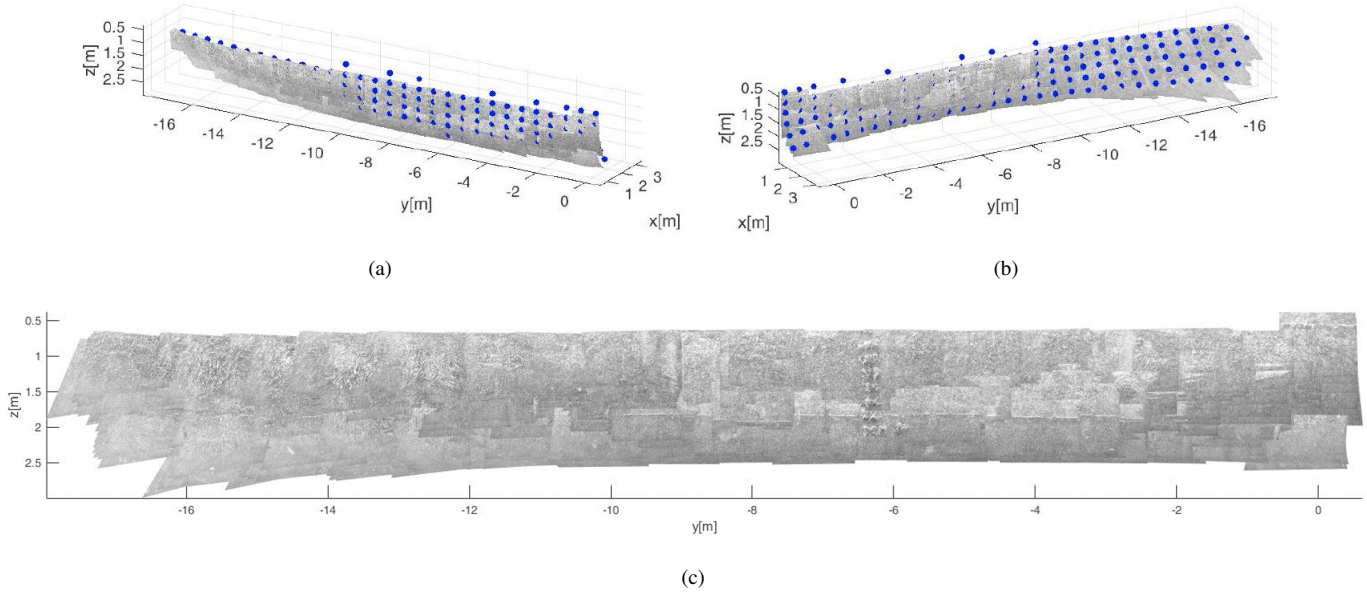


Fig. 9. Reconstruction of the hull surface using plane patches based on the estimated relative pose of the camera toward the hull surface.

5. CONCLUSION

In this paper, a stereo vision system for estimating the relative pose toward the underwater hull surface with consideration of its uncertainty model is proposed. The stereo correspondences are matched using the SURF operator, and 3D point cloud data are extracted based on the geometrical relationship between the 2D feature points and the corresponding 3D points. Assuming that the local hull surface is planar, RANSAC with the plane model is used to reject outlier points. The least squares method using Mahalanobis distance is applied to the remaining inlier points to estimate the plane parameters. With the estimated plane parameters, the relative pose toward the target surface was calculated using the geometrical relationship between the plane and the relative pose. The uncertainty of the relative pose is propagated considering the location uncertainties of the SURF feature points. The effectiveness of the proposed algorithm is demonstrated through an experimental result performed on an actual ship using an AUV.

REFERENCES

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*, ECCV'06, 404–417. Springer-Verlag, Berlin, Heidelberg.

Chung, D., Hong, S., and Kim, J. (2017). Underwater pose estimation relative to planar hull surface using stereo vision. In *2017 IEEE Underwater Technology (UT)*, 1–4.

Elibol, A., Kim, J., Gracias, N., and García, R. (2014). Efficient image mosaicing for multi-robot visual underwater mapping. *Pattern Recognition Letters*, 46, 20–26.

Fischler, M.A. and Bolles, R.C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6), 381–395.

Haralick, R.M. (2000). Propagating covariance in computer vision. In *Proceedings of the Theoretical Foundations of Computer Vision, TFCV on Performance Characterization in Computer Vision*, 95–114. Kluwer, B.V., Deventer, The Netherlands, The Netherlands.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, 147–151.

Hong, S., Chung, D., and Kim, J. (2017). Development of a hover-capable auv system for automated visual ship-hull inspection and mapping. In *OCEANS 2017 - Anchorage*, 1–5.

Hong, S., Kim, J., Pyo, J., and Yu, S.C. (2016). A robust loop-closure method for visual slam in unstructured seafloor environments. *Autonomous Robots*, 40(6), 1095–1109.

Kazmi, W., Ridao, P., Ribas, D., and Hernandez, E. (2009). Dam wall detection and tracking using a mechanically scanned imaging sonar. In *2009 IEEE International Conference on Robotics and Automation*, 3595–3600.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2), 91–110.

Mahalanobis, P.C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2, 49–55.

Negahdaripour, S. and Firoozfam, P. (2006). An rov stereovision system for ship-hull inspection. *IEEE Journal of Oceanic Engineering*, 31(3), 551–564.

Vaganay, J., Elkins, M.L., Willcox, S., Hover, F.S., Damus, R.S., Desset, S., Morash, J.P., and Polidoro, V.C. (2005). Ship hull inspection by hull-relative navigation and control. In *Proceedings of OCEANS 2005 MTS/IEEE*, 761–766 Vol. 1.

Yadav, G., Maheshwari, S., and Agarwal, A. (2014). Contrast limited adaptive histogram equalization based enhancement for real time video system. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2392–2397.

Zeisl, B., Georgel, P.F., Schweiger, F., Steinbach, E., and Navab, N. (2009). Estimation of location uncertainty for scale invariant feature points. In *Proceedings of the British Machine Vision Conference*, 57.1–57.12. BMVA Press.